Translation on Demand Anytime and Anywhere

Gottfried Zimmermann, Ph.D.
zimmer@trace.wisc.edu

Gregg Vanderheiden, Ph.D.
gv@trace.wisc.edu
Trace R&D Center
5901 Research Park Blvd.
Madison, WI 53719

Abstract

In face-to-face conversations people with functional limitations and their interlocutors may encounter communication problems if there is no assistance available. A similar problem arises when the dialog partner is a computer application or web site rather than a human being and its user interface lacks full accessibility. Whether in face-to-face communications or in front of a computer, an instant translation or interpreter service, available anytime and anywhere, would help to bridge the communication gap between the user with special needs and his communication partner.

This paper describes principles and possible applications of a network-based "translation on demand" service. Hereby it concentrates on services for deaf and hard-of-hearing persons (text-captioning on demand, signing on demand) and a service for blind and visually impaired persons (description on demand).

Introduction

The human society heavily relies on communication. This applies to all kind of conversations, whether it is a personal conversations, a presentation in a business meeting or a human-computer dialog. If a communication partner has a functional limitation that prevents him from gaining full access to the provided information there should be a translation from one communication mode to another. Thus for deaf or hard-of-hearing persons audio content must be translated to text or sign language (and often vice versa). Vision-impaired participants may need a description of a picture or any visual object that is essential in a particular context.

People with functional limitations often experience wide communication gaps in everyday's situations. Consider for example a hearing-impaired participant in a meeting where information is mostly exchanged auditorially. However, a sign language interpreter (or a verbal description for a visual object) is not always available. In the same meeting there might be a blind person who has no access to the visual diagram somebody brought in and which is now in the center of the discussion if nobody describes the diagram verbally as the debate goes on. Both, the deaf and blind participants in this example, are precluded from fully participating in this meeting because of their functional limitations.

A similar problem faced in the emerging "information society" stems from the fact that we are more and more dependent on having unrestricted access to online information provided by a public information network. Guidelines help in creating accessible web sites and application interfaces for cross-disability access. However, not

all web sites and applications may conform to these guidelines for a number of reasons. And there is no "guideline solution" for dynamic content, e.g. live images provided by web cams or live audio streams in distributed collaborative environments.

Translation on Demand

Basically, an "assistant on demand" is an individual who could be called up to assist someone with a disability anytime they required it, but who would not be around the rest of the time. The concept of "translation on demand" provides a network-based translation service available anytime and anywhere and may be human or computer based (or both) [Vanderheiden, 1995] [Vanderheiden, in press]. This paper concentrates on three different types of translation on demand services provided for different user groups: text-captioning on demand, signing on demand and description on demand.

Text-Captioning on Demand

Text-captioning on demand instantly translates speech to text for deaf or hard-of-hearing people.

Today's speech-recognition software achieves reasonable recognition rates only with a restricted vocabulary or with speakers for whom the system was trained before. For many everyday-situations where arbitrary people are talking together this is not applicable. But speech-recognition software still can help in getting an accurate speech-to-text translation if a dedicated speaker repeats everything that has been said by other people. In fact, this technique works well even if the dedicated speaker is in remote location and only connected to the other stakeholders via a wide-area network. The network can convey acoustic information in one direction and text-based information in the other direction. This text is then displayed on a screen, hand-held display or special eye glasses (e.g. the Personal Captioner [PCS]).

Another technique for a text-captioning service is a trained person typing on a stenographic keyboard connected to a computer (e.g. the National Captioning Institute [NCI]). This person could be remotely connected to the customer in the same manner as described above.

Signing on Demand

Signing on demand is a remote service accessed via wide-area network. An audio stream from the point-of-need location can be sent to a remote human sign interpreter. On return, a video stream showing the interpreter's signing is sent over the network and posted on a screen or hand-held device for the hearing-impaired person(s). In addition, a video stream from the location to the remote interpreter is needed for sign-to-speech translation. For this direction the interpreter's speech is transferred to the requesting location by an audio stream. An accompanying video stream from the location to the remote service may also help the sign interpreter in getting valuable context information for the translation service.

There are several reasons why sign language translation may be preferable to text translation for certain situations. 1) Sign language (particularly American Sign Language) can express information that is conveyed by speech but not codable in plain text (e.g. emphasis or timing of spoken words). 2) A deaf or hard-of-hearing person may wish to actively take part in a live spoken conversation using sign rather than text. Thus it is more natural to recognize signing and respond in signs than read text and respond in signs. 3) Moreover not all hearing-impaired people can read and understand English at conversational speeds. As a conclusion, speech-to-sign translation is an appropriate means where information exchange mainly relies on spoken language or where hearing-impaired participants lack sufficient reading skills. Signing on demand should provide both translation directions, speech-to-sign and sign-to-speech.

With emerging speech and image recognition, machine-translation and avatar (computer-generated human-like character) rendering techniques this service might be provided in a fully automatic mode. However it might remain remote because of the sheer need of enormous computing power. Thus an audio stream would be transferred to a

remote service application which in return would send movement commands for a signing avatar being rendered on a screen or hand-held display. In the other direction the video image of a signing person would be analyzed by the remote service. Thus the sign language would be translated to synthesized speech and sent back as an audio stream to the requesting location.

Description on Demand

Description on demand translates visual information into verbal form for blind and visually impaired persons where no other verbal description is available.

This service is provided by a person remotely connected to the requesting person. Therefore a video stream (or image) must be provided to the service showing the visual environment or object to be described. The verbal description delivered by the service personnel is then sent over the network and brought to the requestor by speakers, headphones or earbud.

Variations of Application

There is an almost infinite number of possible applications for the translation on demand service. This section will only briefly mention some highlights in order to describe the potential of this service.

Remote Signing for Public Events

At the SuperComputing 99 conference, which was held at Portland's Oregon Convention Center in November 1999, the Trace R&D Center [Trace] demonstrated the feasibility of real-time sign language translation over a high-speed internet for a wide audience [Barnicle et al., 2000]. The audio stream of the plenary session was sent to the remote Trace Center in Madison, Wisconsin, where human sign interpreters provided an instant translation service for the conference. The signing was captured by video and sent back over Internet II to the conference location where it was rendered on a large screen in front of the room.

In this manner a cost-effective high-quality sign language translation service can be provided anytime and anywhere via a high-bandwidth global network

Text-Captioning via Special Eye Glasses

Special eye glasses with a monitor built-in from Personal Captioning Systems [PCS] provide discreet personal captioning. The text captions are provided through wireless transmission and the words seem to "float" about 18 inches in front of the eye. This system can be used in community and social activity locations where an audio signal is easily available (e.g. theatres, movie theatres, conferences venues).

Moreover, in a hypothetical scenario, a hearing-impaired person could use this system in a more personal manner during conversations or meetings with hearing people. The person could hold a pen-like device with a built-in microphone to wirelessly feed a remote service with the environmental sound [Vanderheiden, 1995]. Any spoken word contained in the sound would then be translated into text and instantly shown on his eye glasses. This device would provide individuals who are deaf with the ability to carry on face-to-face conversation with anyone else who might be talking to them. However this scenario presumes that the hearing-impaired person himself has the capability to speak.

As of September 2000, a new service called "Instant Captioning" was announced by Ultratec. Current, in field testing, this will provide a text translation capability using different technology formats [Ultratec].

Remote Description via Earbud

For blind or visually impaired people it would be a great help if they could call for translation on demand anytime

and anywhere. In a hypothetical scenario, if a blind individual became disoriented in an unknown environment, a small wireless camera, built into his cane, could transfer a 360 degree view of the current environment to a description service. A verbal description of the environment would then be sent to the earbud worn by the blind person. Or with the same camera the service could deliver a spoken description of a slide presentation in a business meeting.

Remote Signing in a Collaborative Environment

A collaborative environment is a network of hardware and software systems supporting distributed teams' conversation and collaboration. As an example such a network can facilitate a common meeting of different teams at different locations sharing an electronic white-board and electronic documents.

Collaborative environments are an ideal configuration for a remote sign language translation service. With their network of (wirelessly) connected devices (video cams, large screens, microphones, speakers, electronic white-boards) they offer a complete platform for audio and video capture and rendering. In this scenario the remote location of the sign interpreter is just another node in the collaborative environment and audio and video signals are passed back and forth to this node. The video image of the sign interpreter can be displayed in a section of the large display or, more discreetly, on a personal laptop or hand-held device, wirelessly connected to the underlying network.

This scenario allows a hearing-impaired participant for real-time sign language conversation in both directions. Thus the remote sign interpreter translates other participants' speech to sign and the hearing-impaired participant's signing to speech.

In the future a signing avatar as part of the collaborative environment system could automatically translate spoken content into sign language on demand. Also a video-based sign language recognition system could transform signing from a hearing-impaired participant into audio for hearing participants.

Built-in Translation on Demand for Computational User Interfaces

A functionally impaired computer user from time to time stumbles over (partly or totally) inaccessible user interfaces and web sites. This does not necessarily mean that the developers have been thoughtless or even ignorant. It might also be caused by the mere dynamic nature of the provided content. In this case it would be nice for the user to press a dedicated button (or speak a special command) to launch a text translation, sign translation or video description application. This application would connect his computer to a remote location where the appropriate interpretation service would be provided with or without human assistance.

Thus a text translation service could capture the audio stream of the requesting user's computer and display its content in text form in a separate window. Or the signing service would open a window showing a sign interpreter (or a signing avatar) translating the content of the computer's audio signal into signs. Or the description service would provide a verbal (spoken) description of inaccessible visual content, e.g. diagrams, images and videos.

Built into all browsers and into every operating system's graphical user interface this service button could be the anchor to a powerful "safety net" for unexpected "accessibility crashs". The Trace R&D Center currently works toward such a globally available translation on demand service for the Grid, the next-generation high-speed Internet matrix of services. Being part of the Partnership for Advanced Computational Infrastructure [PACI] funded by the National Science Foundation (NSF), Trace aims to harness today's and tomorrow's high-tech solutions for a globally available translation service on the Grid.

References

[Barnicle et al., 2000] Barnicle, Kitch; Vanderheiden, Gregg; Gilman, Al. "On-Demand" Remote Sign Language Interpretation. In: 9th International World Wide Web Conference, Poster Proceedings, Amsterdam, May 15 - 19,

2000. http://www9.org/final-posters/poster26.html.

[NCI] National Captioning Institute, Inc. http://www.ncicap.org/.

[PACI] Partnership for Advanced Computational Infrastructure, Directorate for Computer and Information Science and Engineering, National Science Foundation.

[PCS] Personal Captioning Systems.

[TRACE] Trace R&D Center, University of Wisconsin-Madison. http://trace.wisc.edu/

[Ultratec] Ultratec, Inc. http://www.ultratec.com/

[Vanderheiden, 1992] Vanderheiden, Gregg. A brief look at technology and mental retardation in the 21st century. In L. Rowitz (Ed.), Mental Retardation in the Year 2000 (pp. 268-278). New York, NY: Springer-Verlag New York, Inc., 1992.

[Vanderheiden, 1995] Vanderheiden, Gregg. Access to global information infrastructure (GII) and next-generation information systems. In A. Weisel, Ed., Proceedings of the 18th International Congress on Education of the Deaf - 1995. Tel Aviv, Israel: Ramot Publications - Tel Aviv University, 1995.

[Vanderheiden, in press] Vanderheiden, Gregg. Telecommunications -- Accessibility and Future Directions. In: Julio Abascal, Colette Nicolle (eds.), Inclusive Guidelines for HCI. Taylor & Francis Ltd., in press.